

Jan/Feb 2005

Genome Technology

Inside Integrated Biology

SMALL MOLECULES GO PUBLIC

**INSIDE:
COMPARATIVE
GENOMICS**

**PROTEIN
FRACTIONATION**

**NIH'S NEW CHEMICAL
GENOMICS INITIATIVE
SENDS RESEARCH
DOWNSTREAM.
HERE'S WHY**

**PLUS:
WHO WILL BENEFIT?
ACADEMICS
AND PHARMA
RESEARCHERS
WEIGH IN**

**NIH's Chris Austin,
Linda Brady, and
James Ingles**



DIVING DEEP INTO THE CHEMICAL GENOME


By John S. MacNeil

Molecular models: Luca Jovine/nuccyl. Cover and Right page photos: Tom Estrin

The NIH Molecular Libraries Initiative attempts to put its own stamp on large-scale small molecule screening

Chris Austin, the co-leader of a new NIH project to build a high-throughput small molecule screening initiative, knows a few things about drug development. After all, the 44-year-old scientist spent six years working for Merck Research Laboratories, where he ran a laboratory focusing on genomic neuroscience and led a team developing new drugs for schizophrenia.

But in his current role as co-director of the Molecular Libraries Initiative, Austin is unlikely to fully exploit his experience working in the drug industry. Although NIH has a mandate to fund and conduct research geared toward improving public health, the institute does not have plans to involve itself in the discovery and development of new drugs — arguably one of the most fundamental approaches to treating disease in modern medicine.



From left:
NHGRI's
Chris Austin,
NIMH's
Linda Brady,
and NHGRI's
James Inglese

To Austin's mind, this is a good thing. The project he leads with Linda Brady of the National Institute of Mental Health is designed to adopt and implement the small molecule (read: drug-like) compound screening technology that the pharma industry has relied on for years to generate lead compounds potentially useful in the clinic. In the hands of NIH researchers and satellite laboratories, however, small molecule screening technology will be put to use almost exclusively as a means of developing new research tools.

"We're not dumb; we're picking the part of drug discovery that's cheap and easy," Austin says. "Making drugs is very much like a Rubik's cube. Most people who pick up a Rubik's cube — most of the time they fail. Drugs are like that. What we're doing is getting the first two squares of the Rubik's cube lined up, and the probability of success of doing that is pretty high." By his calculations, "we're doing two percent of the work, and spending five percent of the time required to make a drug."

Sounds like a good plan, doesn't it? By concentrating their efforts on screening small molecules for activities against novel targets, Austin and Brady hope to discover many new types of molecules that could be used to study biological systems. As another potential benefit, they could jump-start drug discovery in disease areas largely ignored by big pharma. More mainstream drug discovery, the NIH says, will be left to the private sector.

In the best-case scenario for this ambitious project, the next four or five years should prove quite fruitful for public-sector researchers — and even scientists engaged in private-sector drug discovery at biotech and pharma companies. If Austin and his colleagues play their cards right, the project should create a wealth of information on the interactions between small molecule compounds and a wide variety of protein targets. These types of data should prove vital for understanding pathways of interest to both basic biologists and biomedical researchers.

THE FINE PRINT

There are, however, a few details yet to be worked out, and the fine print will determine which groups of researchers will benefit the most, or at all. For one thing, it's widely accepted that big pharma has already mined a significant fraction of the "druggable genome," as the classes of proteins most commonly employed

as drug targets are known, and thus it's more than likely that there will be a fair amount of redundant experiments as NIH scientists and their collaborators seek to duplicate the stores of biochemical know-how already resident in pharmaceutical laboratories. At least initially, then, private-sector drug discovery scientists may not see anything new about the NIH project.

Secondly, the exact contents of the small molecule library NIH hopes to use to screen against large swaths of proteins are still to be determined. To begin with, the MLI will work with a set of compounds fairly standard to pharma labs, but to serve the broadest range of researchers with their investment in chemical genomics, NIH administrators will have to figure out how to adequately sample chemical space in a manner that leads to truly novel findings.

Perhaps most notably, the leaders of the Molecular Libraries Initiative will have to decide how to choose and prioritize the protein targets they hope to screen against the library of small molecule compounds. When the dust settles, it's likely specialists in certain diseases will see results from the screening operation sooner than others.

There's no guarantee that NIH and its collaborators will get it right the first time around — and Austin at NHGRI freely admits there'll be some mistakes during the project's startup period. After all, it is the first time public-sector scientists have

attempted such an extensive small molecule screening project. "Even though we're trying to learn a lot from the private sector, there's the expectation that we'll make a lot of mistakes," Austin says. For this reason, his NIH Chemical Genomics Center will become operational a year before any of the six to 10 pilot screening centers to be funded extramurally (these will be announced in the spring, see sidebar, p. 29). In this way, he says, Austin's team can work out the issues in the logistics of assay acceptance, data deposition, and other operational network issues, while also providing screening capacity to the research community more quickly.

No doubt the challenge of ramping up such a large-scale small molecule screening program is daunting. The numbers speak for themselves: there are theoretically 10^{40} possible variations of small molecule compounds, and NIH expects its library to hold at most 1 million compounds for use in up to



NCBI's Steve Bryant says PubChem will have to be "one size fits all."

500,000 assays. How to cover this range of chemical and biological space — to be precise, finding where they overlap — is no trivial problem.

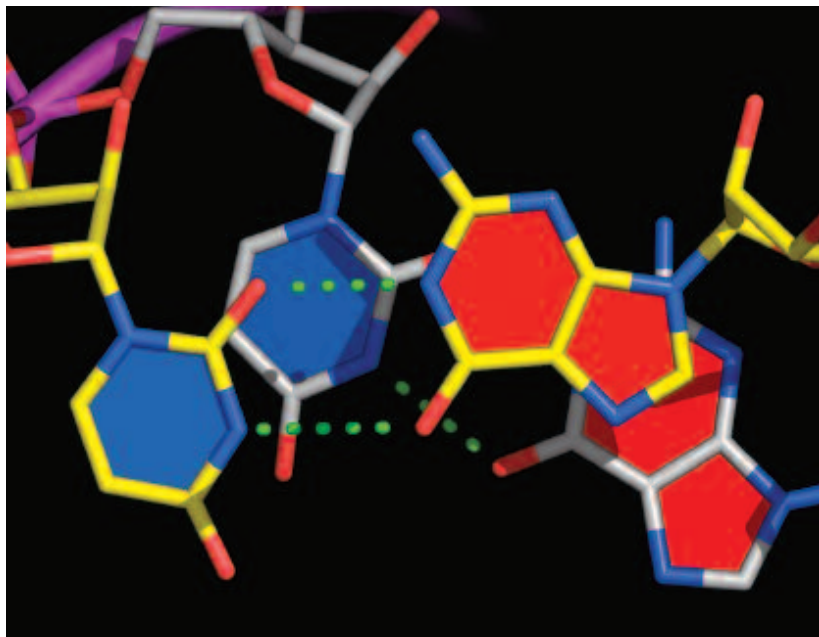
DEALING WITH THE DATA

Perhaps the most well-defined aspect of the Molecular Libraries Initiative at this stage is how the data eventually generated by the project will be presented to the public. PubChem, the data repository for all the structural and interaction information produced by the screening centers, has already made its debut. In October, the National Center for Biotechnology Information launched the prototype version in an attempt to give the database a trial run and familiarize users and potential contributors with its format.

Currently, the database is populated with chemical structures taken from NIST databases and “legacy” data from NIH’s various small-molecule screening projects, such as the National Cancer Institute’s anticancer compound initiative and NIAID’s anti-HIV compound screening data. What makes PubChem different, says Steve Bryant, a senior investigator at NCBI in charge of the database’s implementation, is that the database, as the name implies, is open to the public — one and all.

Given the relatively undefined nature of the data the MLI project will generate, Bryant’s task was to create a database receptive to a wide variety of data formats, ranging from compound concentrations required to inhibit a specific protein to image-processing data. Designing such a “one size fits all” data format required Bryant’s team to create a two-tiered system in which a text field describing the results of a particular assay sits atop the data field containing the actual numerical results of the assay. These fields are then linked to structural data on the compounds screened in the assay, as well as literature references providing additional information when available.

Users access PubChem through the NCBI’s Entrez site, the portal to PubMed and GenBank, among other biomedical search engines. The database is searchable by the various names of a particular compound, authors, or terms used to describe a particular assay, as well as by chemical structure using either the chemical formula or a string notation known as SMILES. Further development efforts include



How will NIH choose the pilot screening centers?

The applications are in, and competitors await the judges’ decision, expected to be released this spring: whom will NIH choose to run the six to 10 pilot screening centers slated to screen NIH’s library of small molecules as part of its Molecular Libraries Initiative?

One thing is clear: the centers won’t be chosen solely on the basis of the lead investigators’ expertise in a particular disease area. According to Linda Brady, a neuroscientist and neuropharmacologist at NIMH and co-leader of the Molecular Libraries Initiative, NIH isn’t necessarily looking for investigators currently running a small molecule screening lab either. The RFA solicited proposals from academic groups interested in developing or expanding their capabilities in assays, screens, and synthetic chemistry operations, as well as from existing groups with established capabilities in these areas. The most important factor, Brady says, is developing a network with a diverse range of high-throughput screening technologies that can be applied to a broad array of biological assays.

Of the 38 groups that applied to operate a pilot screening center, the chosen few will either currently be running high-throughput screens or have had experience running high-throughput screens in the past in the private sector, Brady says. Even private-sector labs are in the running for the grants, which will divvy up \$20 million among six to 10 centers in the first year of the pilot center program.

NIH is initially according the extramural centers only pilot status because of the fundamentally new nature of the project, says Chris Austin, the other co-director of the Molecular Libraries Initiative. For the first three years the centers will operate with relatively small budgets, enough to get the facilities up and running and build their capacity each year. After the third year, it is expected that NIH will “re-compete,” or solicit and evaluate new applications for funding, and choose a smaller number of centers with expanded capacity and larger budgets. This model is similar to that used to build the Human Genome Project Sequencing Consortium, which was similarly new when it began.

— JSM



What Makes Small Molecules Good Research Tools?

Everyone knows that small molecules — as a general class of compounds — can be effective as drugs, but what makes NIH think that small molecule compounds will necessarily be good research tools to study gene and protein function?

In contrast to antisense or siRNA reagents, which block the function of mRNA, a small molecule is designed to interfere with (or promote) the function of the protein itself. Given the multiple splice forms of a particular gene, it's much more effective to design a small molecule to deal directly with the protein, says Chris Austin, a co-leader of the Molecular Libraries Initiative at the National Human Genome Research Institute. "Most of the physiology and biology acts at the protein level," he says. "So it's important to have a tool that manipulates at the level of the protein."

In addition, small molecules are more flexible than other research tools in how they modulate gene function. With a reagent that acts on the mRNA level, the effect is, for the most part, either on or off. A small molecule, on the other hand, can cause much more subtle physiological effects, says Austin. A compound that functions as an allosteric antagonist, for example, acts as a dimmer switch on the target protein's function, whereas an inverse agonist inhibits the target regardless of the presence of that protein's natural ligand. "It's as if small molecules come in cappuccino, latté, all kinds of flavors," Austin says.

— JSM

populating PubChem with structural information obtained from chemical vendor catalogs, adding links to third-party chemical abstracts services that would provide additional pharmacological data for a fee, and instituting an automated data deposition function, says Bryant.

BUILDING THE COMPOUND LIBRARY

In deciding how to construct the MLI's library of small molecule compounds, Austin and Brady are starting small. Initially, it will resemble most standard pharma-scale compound libraries, according to Doug Livingston, a senior vice president for chemistry at Discovery Partners International. DPI, a San Diego-based contract research organization with small molecule synthesis facilities in San Francisco and a proprietary small molecule library in Basel, Switzerland, was awarded the four-year, \$24 million contract to produce, store, and manage the MLI's compound library of up to 1 million chemicals. At the outset, DPI will rely on external vendors, such as Sigma Aldrich, to supply the bulk of the compounds in the library's "base set" rather than synthesize the compounds themselves.

At the moment, Austin and Brady, with the help of an advisory panel assembled from industry and academia, are still determining exactly how to populate the library beyond the base set. Brady, a neuroscientist at NIMH, hopes to see the initiative stimulate the development of chemical tools for use in biological research and in early stage drug development for rare or underserved disease areas such as spinal muscular dystrophy or drug addiction. She envisions the compound library

including the entire set of FDA-approved drugs (active ingredients), a selection of compounds derived from natural products, and sets of compounds known to interact with traditional drug targets such as GPCRs, kinases, and ion channel proteins. Beyond that, Brady and her team are still in the process of defining the parameters they will use to determine which new types of small molecules, known as the "diversity set," should augment the base set. These parameters, Brady adds, include such factors as degree of aqueous solubility, molecular weight, stability, and exclusion of compounds with reactive groups.

"There's been a lot of discussion around how to ensure we're getting new things into the compound library," says Austin at NHGRI. "We're making tools, not just drugs, so we don't want the library to look just like [a pharma compound library]," but some part of the assays would have to be similar to those in pharma in order to establish an equivalent foundation of knowledge on the interactions between drugs and targets, he says. "We're really trying to walk a fine line."

Efforts to explore new areas of biologically active chemical space are not altogether new to academic researchers. Stuart Schreiber at Harvard's Institute for Chemistry and Cell Biology has the best-known research program focusing on synthesizing new chemicals and investigating their role in biological systems. Given Schreiber's success in applying this approach to creating new biological knowledge, NIH's support for this strategy bodes well for the success of its own initiatives. "What Stu Schreiber has done is very inspiring; it's fair to say his publications and accomplishments have stirred

up additional enthusiasm to get access to these kinds of technologies,” says NHGRI Director Francis Collins. “[The MLI] is an effort to expand that capability to make possible greater access to small molecules to answer biological questions. You can’t do it for everyone!”

In November, NIH released two RFAs designed to encourage the development of novel chemical compounds derived from both combinatorial chemistry/diversity-oriented synthesis, targeted synthesis of specific types of compounds, and the isolation and purification of bioactive compounds from natural sources such as microorganisms, marine organisms, or plants. One of the RFAs, designed to create pilot-scale libraries, commits a total of \$3.5 million to fund eight to 12 grants; the other provides \$3 million to fund eight to 10 grants targeted specifically at promoting new methods for isolating natural products useful as biologically-active compounds.

Already, Austin says he expects the MLI compound library will differ from those in pharma by including metabolic intermediates and other established bioactive compounds, as well as known toxic compounds. Most of these compounds are not in pharma collections since they cannot be patented or are not medically attractive. However, they’re perfect for NIH’s purposes because screening such compounds will establish new activities for known compounds, and thus help make connections between different parts of biological space. Furthermore, he says the MLI library will have to diverge from a typical pharma model if NIH researchers hope to find small molecules for new types of assays. For example, he says, he hopes the initiative will identify compounds able to disrupt every protein-protein interaction in a predictable way.

“I don’t think there’s any way we can say now how many compounds we need, or [whether] the kinds of small molecules you need to reach novel parts of genome space are really fundamentally different from the kind of molecules we have now,” Austin says. “What goes into the collection will change over time, very clearly.”

FIRST UP: TARGETS

So to a certain degree the content of the library will be deter-

mined by the types of targets the MLI program decides to study, and in what order. And here’s where it gets murky. Implicit in this prioritization process is the determination of which groups of scientists will benefit first from the data the MLI project produces, and to what extent.

In explaining the rationale for MLI, Austin likes to fall back on the Human Genome Project as both model and inspiration. Like the Human Genome Project, Austin sees the MLI as an effort that relies on a network of technology-intensive laboratories to build easily distributable enabling tools with a genome-wide focus. To this end, he expects each of the six-odd extramural pilot centers to specialize in a particular type of assay or assay technology. Much as the Baylor College of Medicine specialized in sequencing on ABI 3700s, then, the individual pilot centers will concentrate, for example, on developing expertise in yeast-based or whole-organism-based assays.

Where the analogy to the Human Genome Project breaks down is in determining which specific targets to study in an assay. Unlike the HGP, in which individual sequencing centers worked on a particular chromosome or chromosomes for the duration of the project, NIH will select and assign small molecule screening assays to the pilot centers after an ongoing peer review process. The assay solicitation, which NIH will publish early next

year, will ask researchers to submit proposals that target genes, proteins, and cellular/organismic phenotypes associated with any part of the genome or disease.

At this early stage in writing the RFA, Austin adds, the criteria for selecting a proposed assay involve basic biological interest, whether there are existing small molecule probes that can achieve the desired function, and whether the assay is tractable via high-throughput screening. Most of the assays will have to be compatible with 1,536-well format, he says, and will require narrow coefficients of variability.

While Linda Brady’s own research interests center around assays relevant to neurological diseases, she says the reviewers will have a trans-NIH mandate. “The goal is to solicit a variety of innovative biological, biophysical, and cell-based assays for biological targets or processes for which an inadequate array of selective and potent small molecule modulators are available to the public,” she says.

*Like the Human Genome Project,
Austin sees the MLI as an effort that
relies on a network of technology-
intensive laboratories to build
easily distributable enabling tools
with a genome-wide focus.*

In fact, Austin expects the range of proposed assays to mirror the range of interests represented at the 27 institutes and centers that make up NIH, since the small molecule project as a whole is a trans-NIH roadmap initiative. The first indication of this came in the review of the MLI assay technology development grants funded in FY 2004, in which the distribution of biological/disease targets of the assays approximated the relative sizes of the NIH institutes' budgets.

Austin also stresses that the long-range plan for MLI involves a particular emphasis on screening targets out of favor in the pharmaceutical industry. Because of its desire to rapidly create products, pharma tends to focus only on protein targets known to be amenable to interactions with small molecules — the so-called “druggable genome” that includes traditionally receptive targets like GPCRs, ion channels, and kinases. Austin points out that these targets represent only about 3,000 to 5,000 genes, or around 10 percent of the genome, leaving a good 90 percent wide open for exploration with small molecules useful in a capacity as probes or potential therapeutics.

“We’re assuming that there will be a variety of assay types at the different centers. They’ll all have their different flavors,” Austin says. “Together they’ll comprise a true network, and be able to cover as much of the genome as possible.” Austin adds that the MLI could encompass at least a half million distinct assays, through which researchers could study whether particular small molecules promote a protein’s function, interfere with the protein’s function, or otherwise modulate a protein’s function.

That, however, still leaves open the question of what Austin and Brady will decide to highlight during the year or so in which the NIH screening lab will be in operation before the pilot screening centers are up to speed. At this point, Austin says his group is soliciting proposals from within NIH that are designed to validate the small molecule screening technology and test the performance of various assay types. These “guinea pig” assays will include cell-based assays with various forms of read-outs, such as fluorescent, enzymatic, or luminescent technologies, he adds.

What has been set in stone is the laboratory equipment to be installed at NIH for carrying out the screens. In June of last year, NIH awarded the four-year, \$30 million contract to Kalypsys, a San Diego-based manufacturer of robotics for drug discovery. The system, which the NIH Chemical Genomics Center expects to receive sometime this spring, uses robotic liquid handling to dispense targets and small

molecules into individual wells on 1,536-well plates. Kalypsys’ system can accommodate a variety of readout technologies and assays, including whole-organism screens on yeast or zebrafish, and can screen up to 1.5 million small molecules a day, says Kalypsys President and CSO John McKearn. James Inglese, a former Merck researcher serving as director of bio-molecular screening and profiling at the center, will oversee the day-to-day operations of the Kalypsys instrumentation.



Ultimately, however, most researchers familiar with the Mol-



Kalypsys' John McKearn with the robotics his company will install at the NIH Chemical Genomics Center

ecular Libraries Initiative — even those in pharma — agree that the data the project is expected to produce will be quite valuable. The desire to comprehensively explore small molecule interactions with largely uncharted regions of the genome lies in stark contrast to screening efforts in the drug industry, as well as the small-scale screening operations that currently exist in academia, but the ambitious nature of the project should create significant upsides despite the near-term confusion over prioritizing the initial assays.

Nor is NIH alone in trying to create programs that explore the potential of using small molecule/protein target interac-

tions to jump-start research that may rapidly lead to new treatments and therapeutics. The pharma industry has begun to orient itself more toward late-stage drug development, leaving a “chaperone gap” between the discovery of a hit compound and its introduction as an Investigational New Drug, says Ted Spack, senior director of the PharmaSTART program, an effort between SRI International and several universities on the West Coast to advise academic researchers on how to optimize the chances of their discoveries entering the clinic. “It’s a good sign that NIH is facilitating this process,” Spack says.  

NIH: No Designs on Drug Discovery; Private Sector Wary

NIH officials are making a concerted effort to deny that the intended goal of the Molecular Libraries Initiative is drug discovery, but a closer analysis reveals that things aren’t quite so black and white. Perhaps the more important question is, why shouldn’t NIH get involved in drug discovery?

NHGRI Director Francis Collins has a ready answer to that question. First of all, he says, there’s no reason for NIH to upset the well-established relationship between government-funded researchers and the private-sector drug discovery and development industry. Secondly, NIH does not have the expertise to complete all the tasks necessary to bring a drug to maturity, such as toxicology and ADME tests, not to mention clinical trials. Simply put, “NIH couldn’t possibly afford it,” Collins says. “We’re trying to do the part at the very front end, primarily to find research tools,” he adds. “If, in so doing, we find additional molecules [that the drug industry can take forward]: Hooray! That’s a wonderful outcome.”

But Collins and other NIH officials do note one exception to this rule: In the case of rare diseases where the small number of patients discourages the drug industry from investing heavily in finding new therapeutics, NIH does plan to actively seek out small molecules that would be effective treatments. This effort would be restricted to rare diseases and diseases that primarily affect the developing world, where the market is too small to spark pharma’s interest, says NHGRI’s Chris Austin, co-leader of the Molecular Libraries Initiative.

In those instances when NIH does try to jump-start drug development, what’s the strategy for keeping the data pre-competitive? Would public-sector efforts to discover new therapeutics — even for rare diseases — produce data that would be useless to pharma simply because it was public?

“What NIH would do would fall well short of any IP production,” says Collins

Drug discovery is extremely competitive, after all, and from a pharma perspective it might make more sense to go after a small molecule unreported in the literature for which one company could get a head start on its rivals. Again, NIH has a well-practiced answer. “We’ve been very reassured by all observers that what NIH would do would fall well short of any IP production,” says Collins. “If a company got interested [in a particular small molecule], it would undoubtedly need to make many modifications to the compound before they could file any IP.”

Yet some scientists have their reservations, particularly with the idea that MLI’s efforts will necessarily do more than duplicate pharma’s efforts in the public sector. Alan Binnie, a program coordinator at Aventis’ Drug Innovation and Approval Combinatorial Technology Center in Tucson, Ariz., says he’s wary that the NIH effort will devote more effort to drug discovery than it’s letting on. “When they say small molecules, I’m a little bit suspicious because that sounds like drug discovery, and after all, advancing human health is their job,” he says.

René Amstutz, Global Head of Discovery Technologies for the Novartis Institutes of BioMedical Research in Basel, adds that he worries that the quality of the NIH’s small molecules screens will prove secondary to the effort to generate a large and diverse library and run a great number of screens. “It’s an interesting exercise; my personal thinking is that quality issues might lead to misleading data.”

Collins, however, is convinced that the Molecular Libraries Initiative will prove beneficial, even to pharma. “Everybody agrees that the drug industry needs more drugs in its pipeline,” he says. “What we’re hoping to do is make that more productive. I would think that that would lead to greater profits, not less.” — JSM